

# EOSC-Pillar

*Coordination and Harmonisation of National & Thematic Initiatives to support EOSC*

## Cas d'utilisation n°2 « Sciences de l'environnement et de la Terre »

IFREMER/coordination Tâche 6.2  
& partenaires

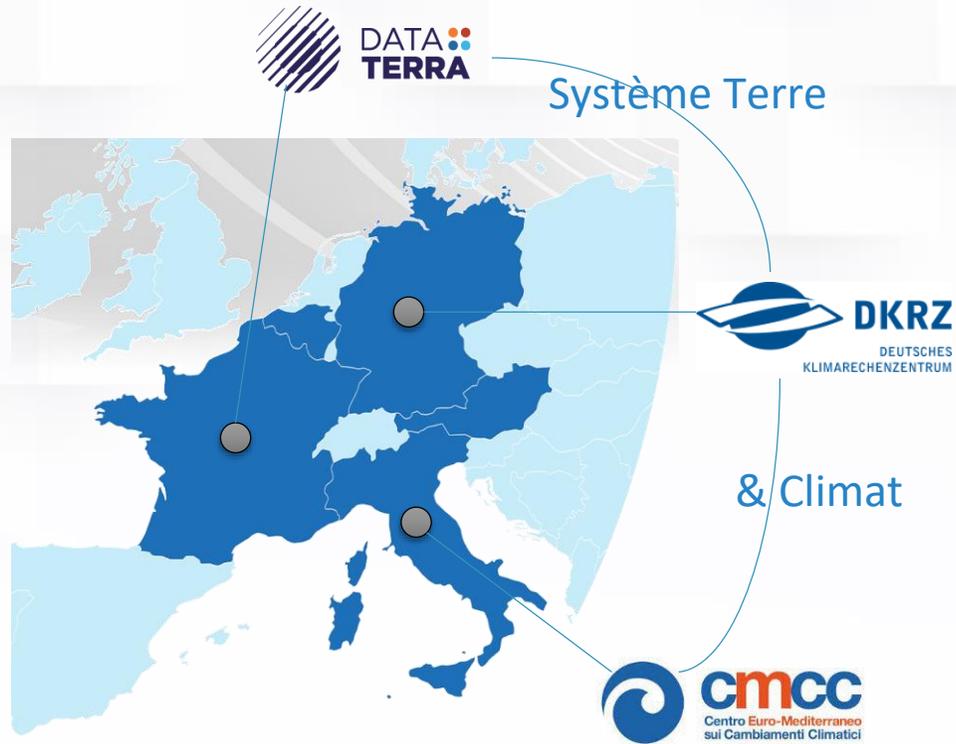
22 Mars 2021 - Journée EOSC-Pillar France



EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.  
This material by the EOSC-Pillar is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



## Infrastructures et Centres de Recherche



## Partenaires

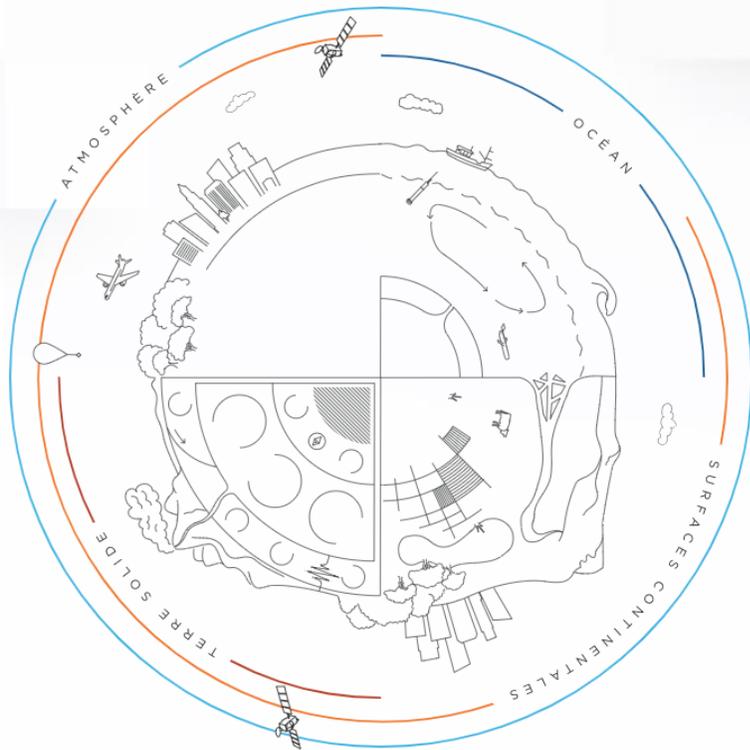
A collection of partner logos arranged in a grid within a blue-bordered box:

- Row 1: French flag, Ifremer logo, cnrs logo, INES logo (Centre Informatique National de l'Enseignement Supérieur).
- Row 2: German flag, DKRZ DEUTSCHES KLIMARECHENZENTRUM logo.
- Row 3: Italian flag, cmcc logo (Centro Euro-Mediterraneo sui Cambiamenti Climatici), INFN logo.

# IR Data Terra au niveau national Français

Mission : organiser de manière intégrée l'accès aux données, en mettant à disposition les données, les produits et des services relatifs à l'observation du système Terre

## 4 compartiments



## 4 pôles de données

Infographic showing the four data poles under the DATA TERRA umbrella. At the top, the DATA TERRA logo is accompanied by the text: 'Infrastructure de Recherche UMS CPST data-terra.org'. Below this, four quadrants represent the data poles, each with a logo, name, and website:

- ODATIS**: Pôle OCEAN, [odatis-ocean.fr](http://odatis-ocean.fr)
- AERIS**: Pôle ATMOSPHERE, [aeris-data.fr](http://aeris-data.fr)
- ForM@Ter**: Pôle Terre Solide, [poletterresolide.fr](http://poletterresolide.fr)
- Theia**: Pôle Surfaces continentales, [theia-land.fr](http://theia-land.fr)

A central image of Earth with a grid overlay is positioned in the center of the four quadrants.

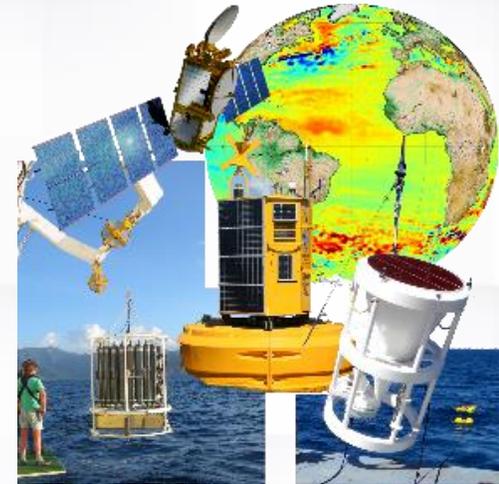
# Le contexte du cas d'utilisation : les besoins

## Sciences des données pour les communautés de “l’environnement et de la terre”

- \* Utilisateurs scientifiques ou analystes de données
- \* **Inter-domaines**, indépendamment des thèmes scientifiques ou des organisations
  - \* Intégration des observations satellite et in-situ
  - \* Multi ou inter-disciplinaires dans les communautés des “sciences de l’environnement et de la Terre”
    - \* Océan / Atmosphère / Surfaces continentales : vent, vagues, chaleur, hydrosphère dans son ensemble, aérosols...
    - \* Océan / Surfaces continentales : débits des rivières, durabilité et sécurité du littoral
    - \* Terre solide Océan / Atmosphère / Océan / Surfaces continentales : Volcans, tremblements de terre, stabilité des pentes, ...
    - \* Climat
  - \* Ou avec d’autres communautés
    - \* Biosphère : Cycle du carbone, climat et impact des changements sur la biodiversité, les habitats, ...
    - \* Activités humaines : artificialisation des sols, impacts, ...

# Le contexte du cas d'utilisation : les données

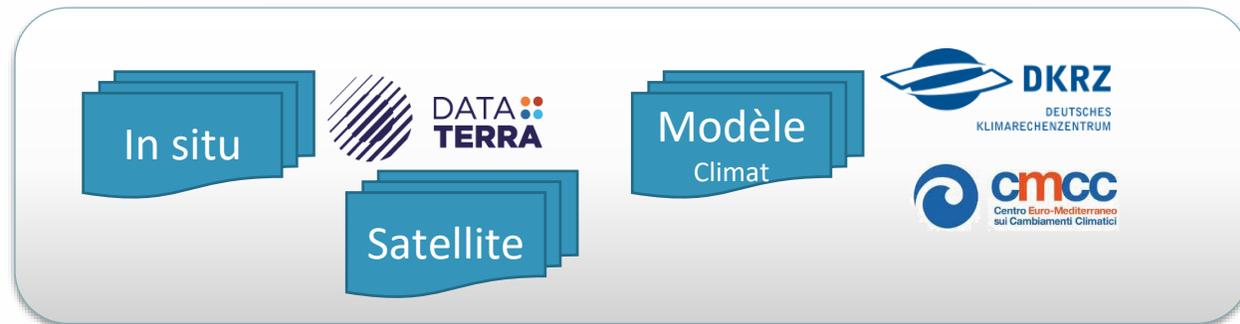
- \* Large panel de sources
  - \* satellite, in situ et modélisation
- \* Volumes importants
  - \* Plusieurs centaines de pétaoctets pour les données satellitaires et les sorties de modèles
  - \* Plus d'un milliard d'observations in-situ (pas extrêmement important mais très nombreuses)
- \* Diversité des types



=> Avec des métadonnées communes spécifiques

- \* localisation sur la terre (latitude, longitude, élévation/profondeur),
- \* temps

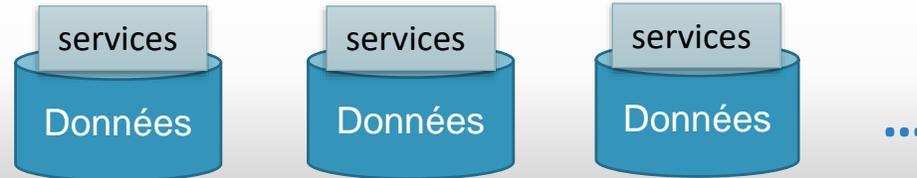
=> Entrepôts de données (& services) multiples et distribués



# Accès aux données et traitement : aujourd'hui



## Entrepôts de données distribués



Scientifique  
ou analyste de données



- Satellite X in-situ
- Harmonisation/agrégation de séries temporelles
- ...

- 1 **Trouver & télécharger** données disponibles & d'intérêt
- 2 **Préparer** rendre les jeux de données (inter)opérables
- 3 **Traiter/analyser** sur moyens locaux 

### Fastidieux

- multiples Portails/catalogues de découverte
- multiples services
- Téléchargement local des jeux de données d'intérêt

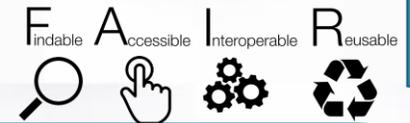
### Nécessite du temps

- Pour comprendre les formats de données
- Pour réaliser l'assemblage

### Limité aux moyens disponibles

- Stockage
- Moyens de calculs

# Accès aux données et traitement : demain

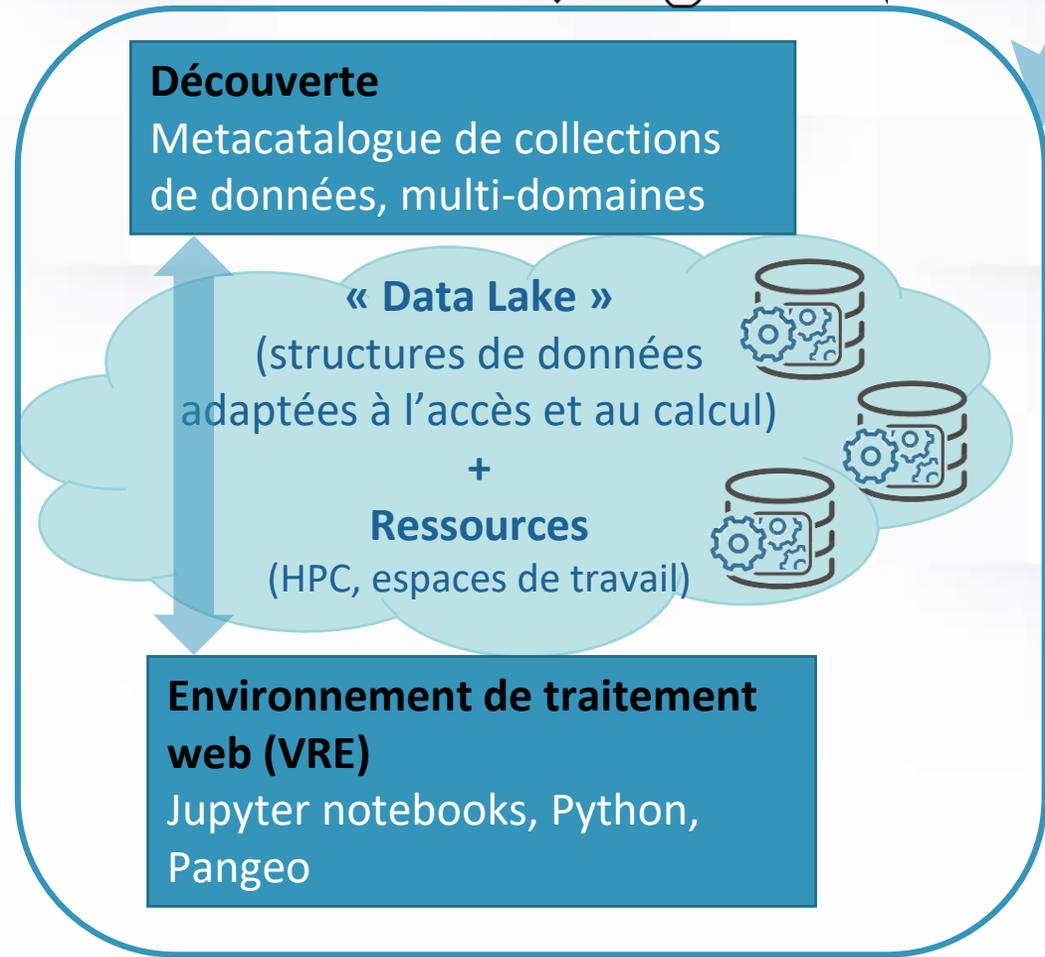
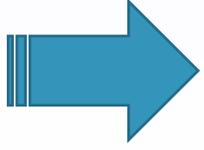
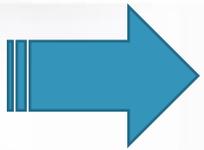


Scientifique  
ou analyste de données



- Traitement Big Data
- Télécharger les résultats pas les données
- Accès avancés aux données

- 1 **Trouver & télécharger** données disponibles & d'intérêt
- 2 **Préparer** rendre les données (inter)opérables
- 3 **Traiter/analyser** sur moyens locaux **distants**



# UC6.2: POC de services avancés "données & traitement"

Démonstrateur de ce que l'EOSC pourrait offrir à la communauté

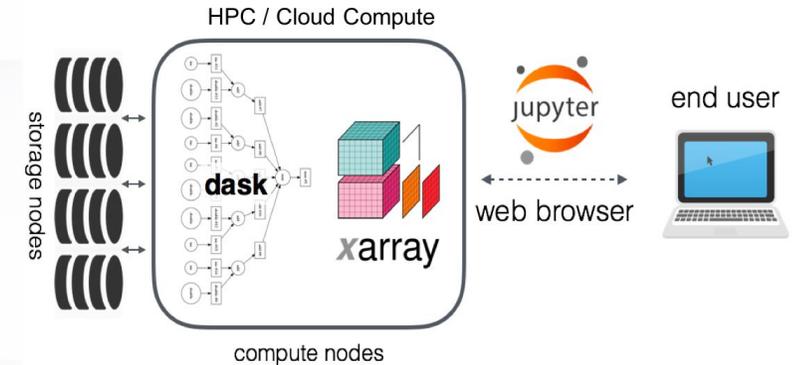
une plateforme « cloud » pour le stockage et le calcul, comme alternative en Europe à Google, permettant aux utilisateurs de conduire des traitements et analyses sur des données volumineuses

Choix de **Pangeo**, plateforme communautaire pour la (géo)science du Big data

- \* Favorise la collaboration autour de l'écosystème scientifique ouvert Python
- \* Implique de nombreuses technologies pertinentes: HPC, containers, notebooks, structures de données avancées ("Data Cubes"), accès distant aux données ...

Découpé en 3 sous use-cases correspond aux 3 types de service

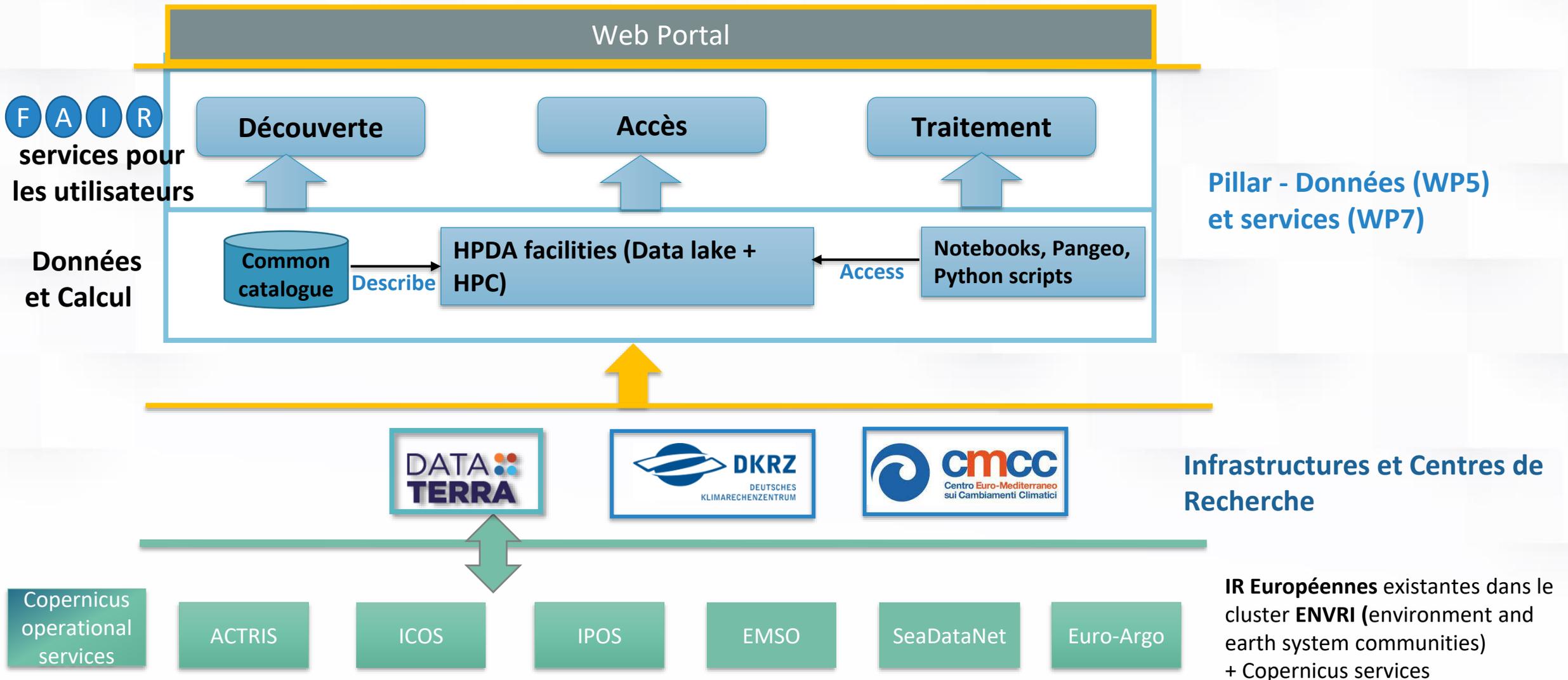
1. **Traitement de type Notebooks "Science des données"**, pour proposer un environnement de traitement web dans l'écosystème Pangeo
2. **Accès**, pour accélérer et faciliter l'accès aux données multi-domaines
3. **Découverte**, à travers l'implémentation d'un métacatalogue inter-domaines



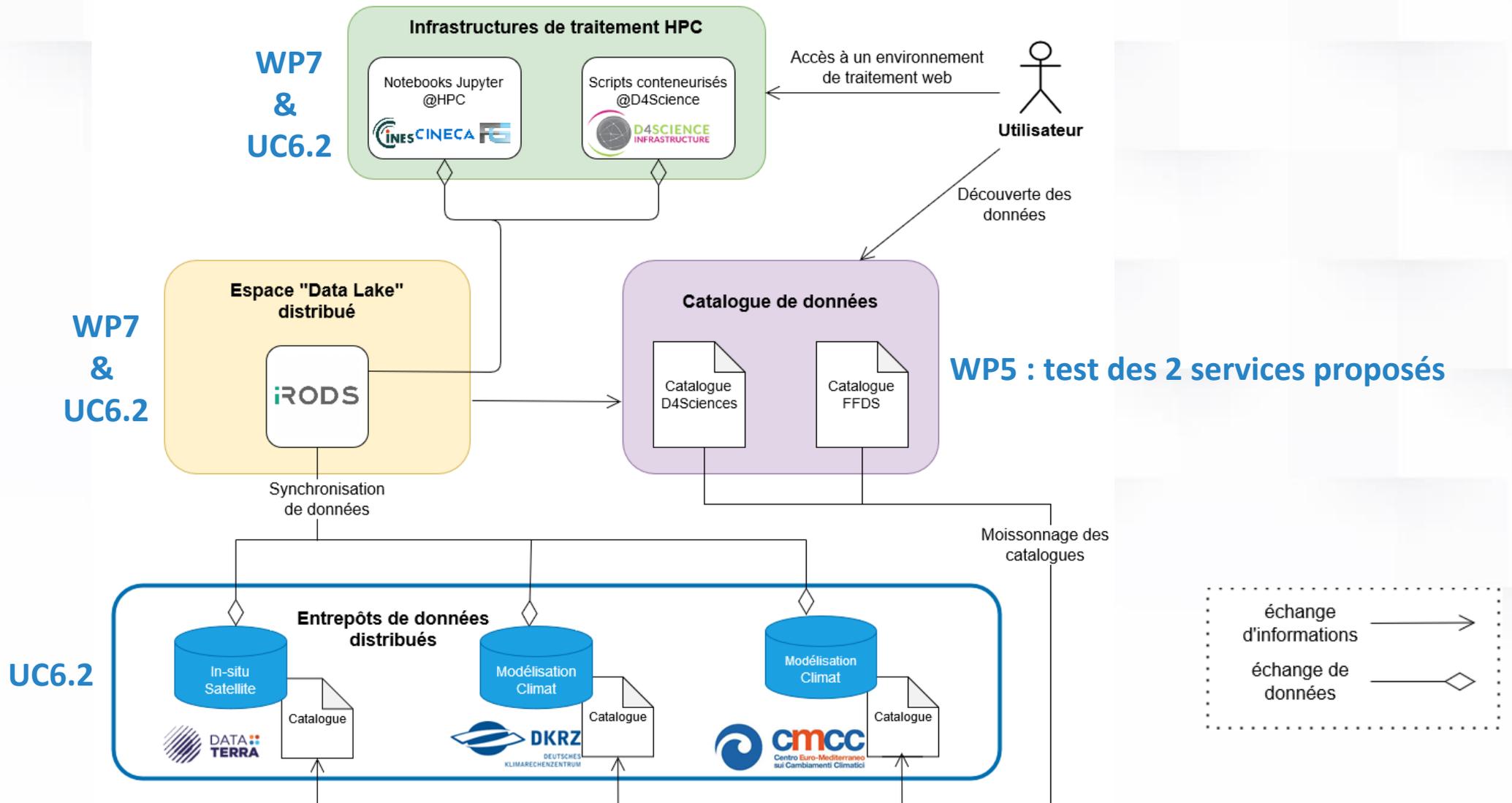
## INTERCHANGEABLE PIECES IN PANGEO (PICK 1 OR MORE FROM EACH ROW)

Data Models	 xarray	 Iris	 pandas
N-D Arrays	 NumPy	 DASK	
Processing Mode	Interactive  jupyter	Batch 	Serverless 
Compute Platform	HPC 	 aws	 Google Cloud Platform
Foundation	 python™		

# UC6.2 : vue d'ensemble



# UC6.2 : Plan de mise en œuvre technique du POC



## UC 6.2 : Feuille de route

- \* **POC en interne Tâche 6.2 par les partenaires: achevé**
  - \* Développement de notebooks « pangeo » exemples pour des collections de données in-situ, satellite et climat
  - \* Choix et validation des formats du « Data Lake »
  - \* Démonstration des notebooks sur les ressources HPDA des partenaires
- \* **POC en collaboration avec les partenaires WP5 et WP7: débuté mi-février**
  - \* Objectif jalon Juin 2021
    - \* Métacatalogue+VRE avec D4Science
    - \* iRODS pour la synchronisation des données vers le « Data Lake »
  - \* D'ici fin 2021 : avancer techniquement le plus loin possible sur le POC dans le cadre de Pillar
    - \* Dont FFDS (sous réserve de méta modèle adapté, GEO DCAT AP)

=> A l'issue du UC6.2 : Recommandations et leçons apprises

# EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

# Merci!

pillar@ifremer.fr

## Get in touch with us!



[www.eosc-pillar.eu](http://www.eosc-pillar.eu)



[@EoscPillar](https://twitter.com/EoscPillar)



[/company/eosc-pillar](https://www.linkedin.com/company/eosc-pillar)



EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650. This material by the EOSC-Pillar Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

